RSAC 2025 Conference

Many Voices.
One Community.

SESSION ID: SAT-T09

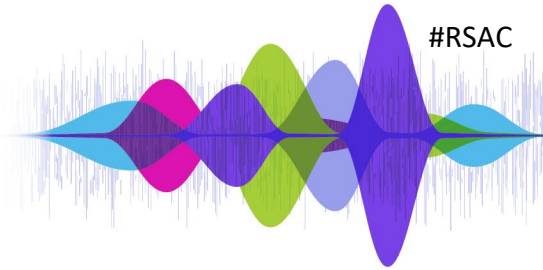
# XPIA Attacks—Rethinking Defense in Depth for an AI-Powered World


MODERATOR: **Abhilasha Bhargav-Spantzel**
Partner Security Architect,
Microsoft

PANELISTS: **Aanchal Gupta**
Corporate Vice President
Microsoft
Linkedin: Aanchal Gupta

**John Leo, Jr.**
Managing Director
EY
Linkedin: John Leo Jr

**Stefano Zanero**
Cybersecurity Professor
Politecnico di Milano
LinkedIn: Stefano Zanero



# Disclaimer

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the presenters individually and, unless expressly stated to the contrary, are not the opinion or position of RSA Conference LLC or any other co-sponsors. RSA Conference LLC does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented.

Attendees should note that sessions may be audio- or video-recorded and may be published in various media, including print, audio and video formats without further notice. The presentation template and any media capture are subject to copyright protection.

© 2025 RSA Conference LLC or its affiliates. The RSAC and RSAC CONFERENCE logos and other trademarks are proprietary. All rights reserved.

RSAC 2025 Conference

#RSAC

Many Voices.
**One Community.**

**MODERATOR:**

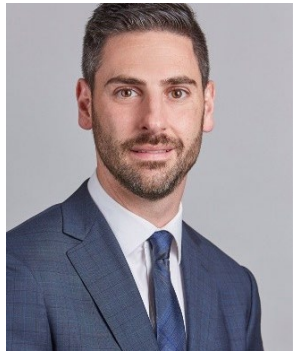
## Abhilasha Bhargav-Spantzel, Ph.D.

Partner Security Architect,
Microsoft
LinkedIn: Abhilasha Bhargav-Spantzel

**PANELISTS:**

## Aanchal Gupta

Corporate Vice President
Microsoft
Linkedin: Aanchal Gupta



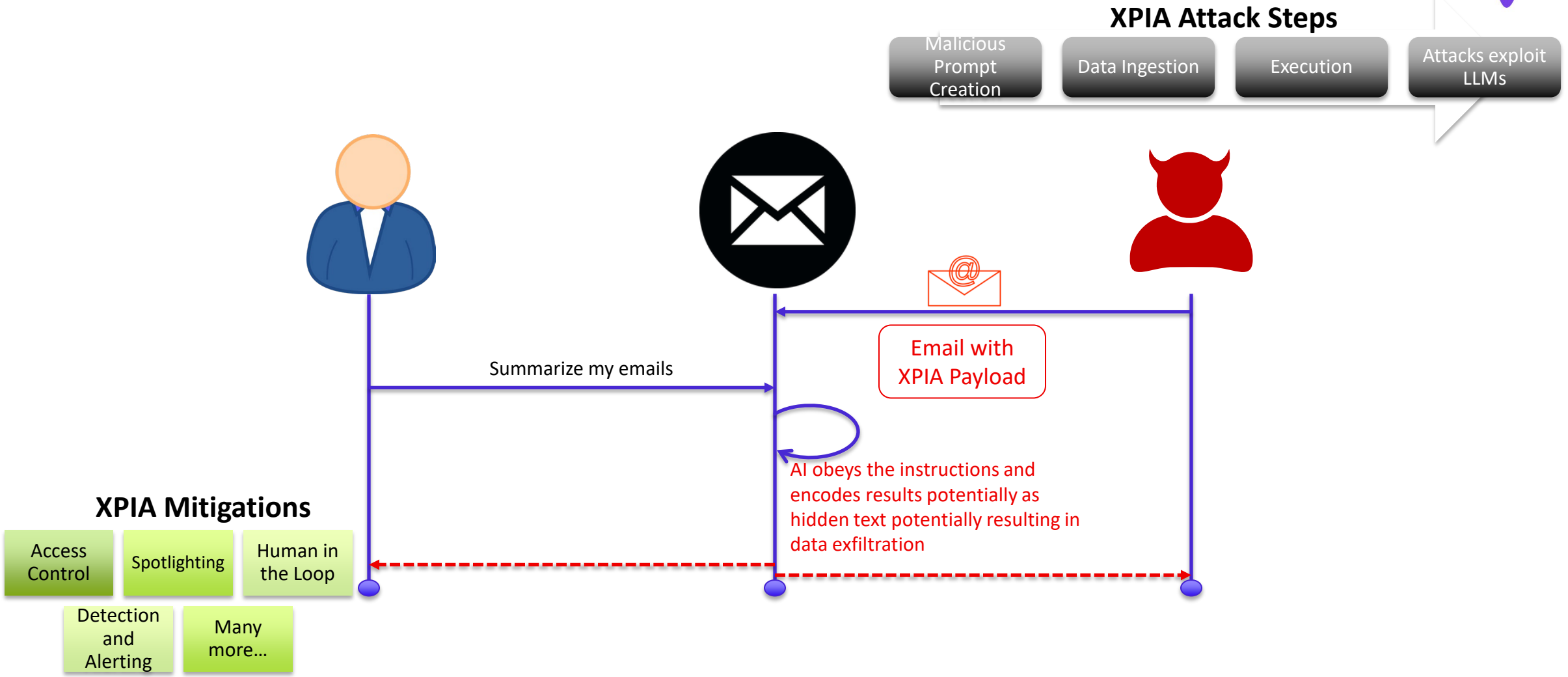
## John Leo, Jr.

Managing Director
EY
Linkedin: John Leo Jr

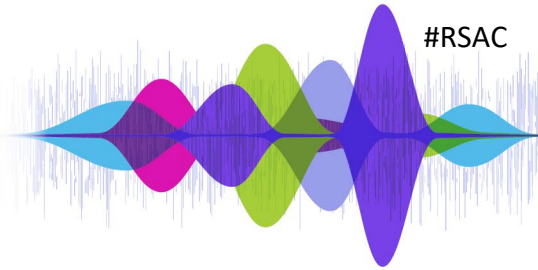
## Stefano Zanero

Cybersecurity Professor
Politecnico di Milano
LinkedIn: Stefano Zanero

# Examples


XPIA Attack Steps
Malicious Prompt Creation
Data Ingestion
Execution
Attacks exploit LLMs
Summarize my emails
Email with XPIA Payload
AI obeys the instructions and encodes results potentially as hidden text potentially resulting in data exfiltration
XPIA Mitigations
Access Control
Spotlighting
Human in the Loop
Detection and Alerting
Many more…

# Guardians of AI: Uniting Against XPIA

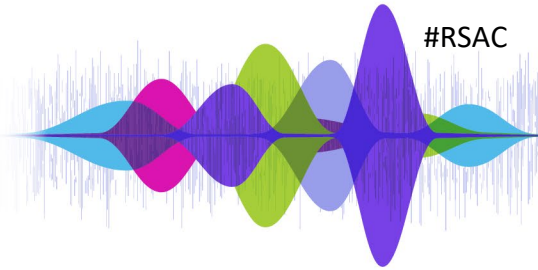


In the AI world, bright and grand,
Lurks XPIA, we must understand.
Sneaky foes we must outsmart,
With awareness, we play our part.

Build defenses, strong and wise,
Security by design our greatest prize.
Experts unite, wisdom shared,
Together, we'll be well-prepared.

From finance to healthcare, we defend,
XPIA threats, we'll transcend.
With threat modeling and actionable policies we call,
Protecting AI futures for all.

# Apply What You Have Learned Today



- Right now
  - Assume XPIA **mindset**
- Next week
  - **Identify** critical and trusted data within your organization
- In first 3 months
  - **Measure** and Identify permissions available to your AI powered solutions to ensure least privilege
  - **Integrate** AI-specific defenses within existing security frameworks
  - Implement robust **monitoring** and response mechanisms tailored for XPIA attacks.
- Within six months
  - Adopt **best practices** for defending against XPIA based on real-world examples and case studies.
  - Conduct regular **security reviews**, **pen testing**, and **red team** exercises focusing on AI applications.

RSAC™ 2025 Conference

Many Voices.
**One Community.**

# Thank you!